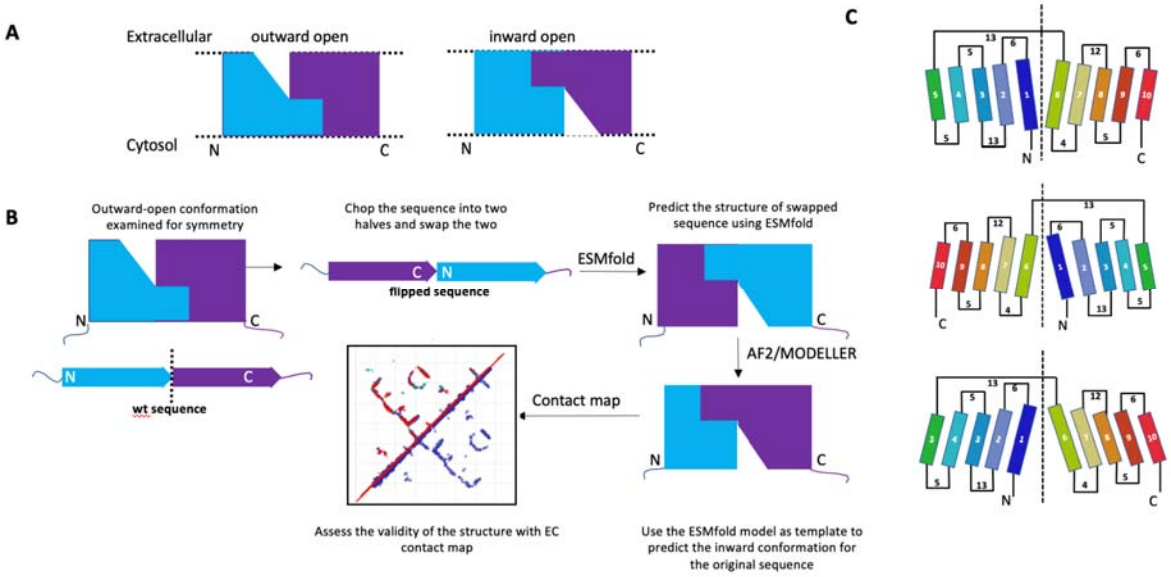
Modeling Alternative Conformational States of Pseudo-Symmetric Solute Carrier Transporters using Methods from Deep Learning

G.V.T. Swapna[1,2], Namita Dube[1], Monica J. Roth[2*], and Gaetano T. Montelione[1,*]

[1]Dept. of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA. [2]Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway NJ 08854 USA.



ABSTRACT

The Solute Carrier (SLC) superfamily of integral membrane proteins function to transport a wide array of small molecules across plasma and organelle membranes. SLC proteins also function as important drug transporters and as viral receptors. Despite being classified as a single superfamily, SLC proteins do not share a single common fold classification; however, most belong to multi-pass transmembrane helical protein fold families. SLC proteins populate different conformational states during the solute transport process, including outward-open, intermediate (occluded), and inward-open conformational states. For some SLC fold families this structural "flipping" corresponds to swapping between conformations of their N-terminal and C-terminal symmetry-related sub-structures. Conventional AlphaFold2, AlphaFold3, or Evolutionary Scale Modeling methods typically generate models for only one of these multiple conformational states of SLC proteins. Several modifications of these AI-based protocols for modeling multiple conformational states of proteins have been described recently. These methods are often impacted by "memorization" of one of the alternative conformational states, and do not always provide both the inward and outward facing conformations of SLC proteins. Here we describe a combined ESM – template-based-modeling process, based on a previously described template-based modeling method that relies on the internal pseudo-symmetry of many SLC proteins, to consistently model alternate conformational states of SLC proteins. We further demonstrate how the resulting multi-state models can be validated experimentally by comparison with sequence-based evolutionary co-variance data (ECs) that encode information about contacts present in the various conformational states adopted by the protein. This simple, rapid, and robust approach for modeling conformational landscapes of pseudo-symmetric SLC proteins is demonstrated for several integral membrane protein transporters, including SLC35F2 the receptor of a feline leukemia virus envelope protein required for viral entry into eukaryotic cells.

*Corresponding authors.

Abbreviations: AF2 – AlphaFold2 Multimer; AF3 – AlphaFold3, EC - Evolutionary Covariance; ESM-Evolutionary-Scale Modeling, LDDT – Local-Distance Difference Test; MD – Molecular Dynamics; ML – Machine Learning; mmCIF - macromolecular Crystallographic Information File; MSA – Multiple Sequence Alignment; PDB - Protein Data Bank; pLDDT - predicted Local-Distance Difference Test, a confidence score predicted from ML, TM – Template Modeling score to assess similarity between two protein structures.

INTRODUCTION

Proteins adopt multiple conformational states which are essential to their functions. While AlphaFold2/3 (AF2/3)[1], Evolutionary Scale Modeling (ESM)[2], and related machine-learning methods[3,4] can provide accurate structural models of proteins, for systems that adopt multiple conformational states conventional AF2/3 and ESM calculations generally identify only one of the multiple states observed experimentally[5-12]. Recently, significant advances have been reported using modified AF2 protocols and enhanced sampling methods to accurately model multiple conformational states of proteins, including integral membrane proteins[13]. Promising approaches use a conventional AF2 platform with curated input such as (i) state-annotated conformational templates[14], (ii) shallow multiple sequence alignments (MSAs) chosen either randomly (AlphaFold-alt)[6,15] or by clustering homologous protein sequences (AF-cluster)[8], (iii) very shallow and even single protein sequences[9,16] that allow knowledge inherent to the AI to dominate the modeling process, or (iv) using MSAs masked at multiple positions, as implemented in both (SPEACH-AF)[17] and AF-sample2[12], to switch the prediction toward alternative conformational states. AF2 calculations using network dropouts (AF-sample) can also generate conformational diversity[18-21]. Despite these advances, however, challenges remain in consistently modeling the multiple alternative conformational states of proteins observed experimentally. In particular, these enhanced sampling methods are only successful for about 50% of experimentally-available alternative conformer pairs[9,11], leading to the suggestion that at least some of the cases of successful modeling of alternative conformational states rely on some kind of memorization by the AI, rather than its inherent "learning" of protein structure principles[9]. To the degree that memorization is required for a particular class of proteins for successful prediction of alternative conformational states, more robust methods leveraging the tools of AI-based modeling are required.

The Solute Carrier (SLC) superfamily of integral membrane proteins function to transport a wide array of solutes across the plasma and organelle membranes. The superfamily includes more than 66 SLC protein families (https://www.bioparadigms.org/slc/intro.htm), each including many individual proteins. SLC proteins transport a wide array of molecules, including sugars, amino acids, vitamins, nucleotides, metals, inorganic ions, organic anions, oligopeptides, and drugs[22-25]. Some are orphan transporters with no known substrate. SLC proteins can also function as receptors for viral entry into the cell[26]. They constitute a major portion of all human transporter-related proteins and play key roles in human health and disease[24,25,27].

Despite being classified as a single superfamily, the various SLC fold families do not share a single common fold classification and are not all phylogenetically related. For example, the two most common SLC fold families, the major facilitator superfamily (MFS) fold, which constitute the largest class of SLC proteins, and the LeuT fold, another important class of SLCs, are topologically and structurally distinct[24]. However, despite these differences, many SLC transporters have a characteristic structural architecture with pseudo two-fold symmetry, where the two halves of the protein structure are related by a two-fold symmetry axis in the plane of the membrane bilayer[24,28]. These halves have a similar fold but non-identical conformations, enabling the protein to adopt multiple conformational states essential for its function. MSF-fold SLC proteins have a "6+6" topology comprised of two "inverted pseudo-repeat" 6-helical bundles with antiparallel orientations related by a pseudosymmetry axis, while the strikingly similar but topologically distinct LeuT-fold membrane proteins feature two 5-helical bundles with "inverted pseudo-repeat" sequences that form structures related to one another by a pseudosymmetry axis[24]. Some (but not all) other SLC proteins also have folds with internal structural pseudosymmetry[24].

SLC proteins populate different conformational states during the transport process, including "outward-open", with a surface cavity directed one way, intermediate states (i.e., occluded, with no surface cavity), and "inward-open" with a surface cavity directed to the opposite side of the membrane[23,24]. These "inward-open" and "outward-open" conformational states are sometimes called inward-facing and outward-facing states in the literature. Crystal structures have been solved for inward-open, occluded, and outward-open states of several MFS and LeuT SLC proteins; for a few SLC proteins both inward and outward-open states have been determined by X-ray crystallography or cryoEM[29-33]. This conformational "flipping" confers an "airlock" or "revolving door" function, which underlies their mechanisms of symporter or antiporter solute transport[23,24,28]. The switch between outward- and inward-open states results from swapping of the conformations of the N-terminal and C-terminal symmetry-related sub-structures, in which the N-terminal helical bundle switches to adopt the conformation of the C-terminal helical bundle, while simultaneously the C-terminal helical bundle switches into the original conformation of the N-terminal helical bundle. These dynamic structural and biophysical properties confer to SLC proteins their functions as gates for symporter and antiporter transport of biochemically-important solutes and biomolecules[24,25].

Both experimental and computational studies of SLC proteins have provided important insights into the role of these conformational dynamics in solute transport. Computational methods have the potential to have a significant impact in understanding structure-function relationships of SLC proteins and to guide the design of experiments. However, as they are medium-sized integral membrane proteins, molecular dynamics simulations are quite challenging, requiring powerful computing resources, accurate potential energy functions, and appropriate simulation of membrane-mimicking environments. The evolving AI-based enhanced sampling methods outline above can sometimes provide models of multiple conformational states of SLC proteins, but are not always successful[6-12]. These observations suggest the need for more robust methods for addressing this important class of membrane protein transporters.

Importantly, multiple conformational state modeling of proteins can be guided by evolutionary covariance (EC) analysis of functionally-preserved direct contacts, which can provide information about contacts present in the two (or more) states adopted by the protein structure [34-40]. Of special significance for SLC proteins, is their unique pseudo-symmetrical transport mechanisms, which provides the basis for classical method of modeling the inward-open (or outward-open) conformations of some SLC proteins from knowledge of their outward (or inward) open conformations by swapping the pseudo-symmetric structures of the N- and C-terminal halves, and then using the resulting virtual structure as a template to model the alternative conformational state[28,41-47]. Although it seems obvious to combine these two concepts and utilize EC-based contact information together with swapping of pseudo-symmetric structures, this approach has not been previously implemented as a general strategy for modeling SLC proteins.

Here we describe a simple and robust approach for modeling alternative conformational states of pseudo-symmetric SLC proteins using a combined ESM – template-based-modeling process inspired by the methods of Forrest and others[28,41-47]. In this approach, templates are generated from a "flipped virtual sequence" using ESMFold[2], and template-based modeling is then performed using either AF2[1] or, where training bias will impact the AF2 structure prediction, with the template-based modeling software MODELLER[48]. First, an ESM-AF2 approach was used to model the inward- / outward-open forms of two SLC proteins, human ZnT8 (SLC30A8, a Zn transporter) and *Escherichia coli* D-galactonate:proton symporter (SLC17, a MFS superfamily transporter) for which experimental structures of both outward- and inward-open states are available, and the resulting models of alternative conformations were validated by comparison against atomic coordinates determined by cryoEM or X-ray crystallography. These models were also validated against EC-based contact maps. For two additional SLC proteins, *Zea mays* CMP-sialic acid transporter 1 (SLC35A1) and *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 (SLC35D subfamily) the outward-open forms are available as experimental structures. As AF modeling was found to be biased towards these states, the alternative inward-open forms were modeled with an ESM-MODELLER process, and then validated by comparison against EC-based contact maps. For SLC35F2, neither inward nor outward-open experimental structures are available. The outward-open form was modeled using conventional AF2, and the inward-open conformational state was then modeled using the ESM-AF2 process. Both the inward- and outward-open structures were then validated against EC-based contact maps. Although bias was observed using the ESM-AF2 process for other SLC proteins for which one of two possible conformational state was available in the PDB, the ESM-MODELLER approach was successful in modeling both inward and outward-facing states of several additional pseudo-symmetric integral membrane proteins, which were validated by comparisons against EC-based contact maps.

METHODS

Evolutionary covariance (EC) - based contact predictions. EC-based contact predictions were performed using evolutionary covariance analysis with *NeBcon* (Neural-network and Bayes-classifier based contact prediction) https://seq2fun.dcmb.med.umich.edu//NeBcon/, a hierarchical algorithm for sequence-based

protein contact map prediction[49], with a probability threshold of 0.7. A second server, *EVcouplings server*[34] https://evcouplings.org/ was also used to confirm these contact predictions.

Contact maps for experimental and predicted structures were obtained from *CMview*[50], an interactive contact map visualization and analysis tool. Contact maps were generated for interresidue Cα distances of < 10.0 Å. The contact lists generated from protein structure models were then imported into excel spreadsheets for overlay and comparison with the EC-based predicted contacts.

AlphaFold2, ESMfold, and MODELLER modeling. AlphaFold2[1] modeling was performed using Colabfold v1.5.5 server[51] with *AlphaFold2.ipynb* scripts. The standard AF2 modeling in this study used no templates, default multiple sequence alignments (MSAs), recycle of 12, and with random dropouts, though other protocols were also assessed. The Amber-relaxed top-ranked model was taken as the final predicted structure. Evolutionary Scale Modeling (ESMfold)[2] models were generated using the *ESMFold_advanced.ipynb* colab script. Models were generated with random masking of input sequences (defined by masking_rate of 0.15), stochastic_mode="LM" that uses no dropout, and recycle of 12. The model with the maximum pTM score was selected as the final model. A locally installed version of MODELLER 10.4[48,52] was used for conventional template-based modeling. 20 models were predicted for each run and the model with the best DOPE (Discrete Optimized Protein Energy score) was selected as the representative structure.

AlphaFold-alt. Enhanced sampling using shallow MSAs with AlphaFold-alt (AF-alt) was carried out as described by Meiler and co-workers[6], using scripts kindly provided by Dr. Davide Sala and executed on a local cluster of 4 A100 Nvidia HGX GPU processors. In each AF-alt run, 480 models were made using randomly-generated shallow MSAs of 16-32 sequences. 30 models were generated for each MSA depth of 16 to 32 sequences. Each run was < 3 hrs. No structural templates were used. For each model, disordered N- and C-terminal regions were removed and the average pLDDT score (<pLDDT>) was then computed for all of the remaining residues.

AF_Sample and AF_Sample2. Massive sampling was carried out using *AF_Sample* and *AF_Sample2* of Wallner and co-workers[12,18,19], executed on a local cluster of 4 A100 Nvidia HGX GPU processors, using protocol details described elsewhere[21]. *AFSample* inferences used various *AF-Multimer* model weights (v2.1.2, v2.2.0, and v2.3.2). In all cases modeling was performed with no templates. When using *AF-Multimer v2.1.2*, modeling was performed using 21 max_recycles, with *v2.2.0* with the default of 3 max_recycles, and with *v2.3.2* using 9 max_recycles. *AF_Sample2* inferences used the same variation in *AF-Multimer* model weights as the *AF_Sample* runs. In all cases inference was run with no templates and 3 max_recycles.

Hydrogen atoms were added to files generated by *AF_Sample*, *AF_Sample2*, and *AF_Alt* using a custom script which employs the Amber force field, analogous to the method employed by the original AF2 manuscript (Jumper et al, 2021). These scripts are provided at

https://github.rpi.edu/RPIBioinformatics/FilteringAF2_scripts. Each of these enhanced sampling methods can be quite aggressive in generating conformational diversity in addition to models that are not physically reasonable: e.g. incorrect amino acid chirality, non-native cis peptide bonds, and other biophysically incorrect features, particularly in the not-well-packed residue segments of the modeled proteins. The most egregious of these physically unreasonable models were identified and removed, as described elsewhere (Spaman et al, manuscript in preparation). The resulting relaxed models were used for further analysis.

Statistical methods. Backbone root-mean-squared deviation (RMSD) and global distance test (GDT) scores for structural comparisons were performed using the methods of Zemla implemented on their public server http://linum.proteinmodel.org/ [53].

Data repository. Key data generated in this study are available at https://github.rpi.edu/RPIBioinformatics/SLCModeling.

RESULTS

The challenge we address arises from the fact that conventional AF modeling will generally provide only one of the multiple conformations of SLC proteins when only one of these states was available as an experimental structure at the time of training[6,7,9-12]. Even enhanced sampling methods successfully generate alternative conformational states for only for some multistate proteins[6,8-12]. These observations motivate the need for robust methods for modeling alternative conformational states (outward-open vs inward-open) of SLC proteins, at the very least for use as reference states for assessing the evolving deep learning methods for generating alternative conformational states of proteins.

ESM-AF2/MODELLER protocol. The ESM-AF2/ ESM-MODELLER process for modeling alternative conformational states of SLC transporters that have structural pseudo-symmetry is outlined in **Figure 1**. It is based conceptually on methods used for other pseudo-symmetric SLC proteins[28,47], in which the pseudo-symmetric halves of the transporter are first identified as an N-terminal protein sequence (blue in **Figure 1**) and C-terminal protein sequence (purple in **Figure 1**), and the N-terminal protein sequence is then modeled using the C-terminal segment as a structural template, and the C-terminal protein sequence is modeled using the N-terminal segment as a structural model. However, application of this method using conventional modeling methods can be challenging if the sequence similarity in these two halves of the protein sequence is low, making it difficult to determine the correct alignment for template-based modeling. In the ESM-AF2/MODELLER process, the N-terminal (blue) and C-terminal (purple) segments of protein sequences are first swapped to create a *virtual flipped sequence*. The entire structure of this virtual sequence is then modeled using *ESMfold*, a large-language model based method that requires no templates and only a single protein sequence. The resulting *virtual protein structure* is then used as a structural template to model the original protein sequence using template-based modeling with AF2 or MODELLER.

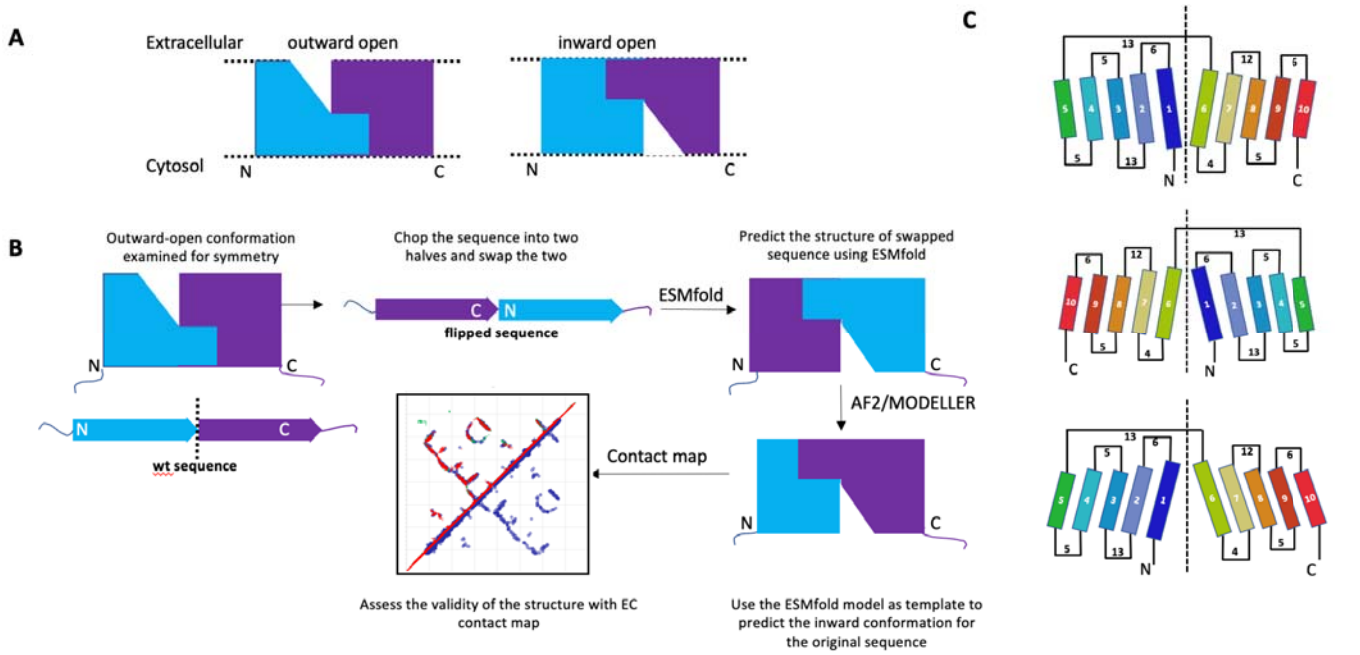


Fig. 1. The ESM-AF2/MODELLER protocol for modeling alternative conformational states of pseudo-symmetric SLC proteins. (A) cartoon representation of inward/outward-open conformers representing the pseudo-symmetry of the helices, with pseudo-symmetry halves indicated in blue and purple. (B) Protocol to model inward/outward-open conformers for symmetric helical transmembrane proteins (C) Topology diagrams showing the conformational flip of a representative 10-helical SLC protein (SLC35F2). The vertical dotted line represents the symmetry axis of the pseudo-symmetric halves of the SLC protein. Numbers represent the number of residues in the membrane-external loops. The top image represents the outward-open state, the middle image is the *ESMfold* virtual protein structure generated from a virtual flipped protein sequence, and the bottom image the inward-open state generated by comparative modeling using the virtual protein structure as a modeling template.

In this study, an *ESMfold* structure model generated from a virtual flipped sequence was used as a custom template for template-based modeling using AF with low MSA (16 - 32), recycle of 12, and with dropout. A shallow MSA is used so that the template information dominates the modeling process. This "comparative modeling" step, using the *ESMfold* model as a template and determining the sequence alignment from structural alignment, was also performed using MODELLER[48,52]. It could also be done using SwissModel[54] or other template-based modeling methods. Finally, the original (e.g., outward-open) and final (e.g., inward-open) structures are validated by comparison against the EC-based contact map that will generally include predicted contacts for both conformational states.

Validating the ESM-AF2 modeling protocol. As an initial test case of the ESM-AF2 method for modeling alternative conformational states of SLC proteins, we selected human ZnT8 (SLC30A8), a 2 x 320-residue homodimeric integral membrane protein Zn-transporter, a representative SLC protein for which structures have determined by cryoEM[30] (PDB ids: 6xpd, 6xpde, and 6xpf, at resolutions of 3.9 Å, 4.1 Å, and 5.1 Å, respectively). ZnT8 (PDB id: 6xpf) has two subunits; in the absence of Zn, chain-A is in an inward-open conformation and chain-B in an outward-open conformation. Conventional AF2-colab

calculations using the standard protocol outline in the Methods section provided a structure with inward-open conformation, in agreement with the cryoEM inward-open structure 6xpf-A (Cα RMSD 2.00 Å). (**Figure 2A**). We then used the ESM-AF2 modeling protocol outlined in **Figure 1** to model the outward-open conformational state, and compared the resulting model with the experimentally-determined outward-open cryoEM structure. The computed outward-open conformation of ZnT8 has excellent agreement with experimental outward-open structure 6xpf-B, with backbone C RMSD of 1.09 Å (**Figure 2B**). We also compared residue-residue contact maps for the experimental and ESM-AF2 outward-open models with each other and with the EC-based contact map generated from multiple-sequence alignments of ZnT8 homologs (**Figure 2C,D**). The computed inward-open structure, modeled with AF2, has a contact map that is nearly identical to that of the experimental inward-open structure (**Figure 2C**); the outward-open structure computed from the inward-open structure using the ESM-AF2 protocol is also essentially identical to the experimental outward-open structure (**Figure 2D**). While many ECs are common to both the outward- and inward-open conformations, the ECs contain information about both states, and several are unique to each conformation; i.e. there are 6 unique ECs for outward-open and 6 unique ECs for inward-open states. These several ECs unique to the outward/inward-open conformations superimpose on top of the corresponding unique contacts in the outward and inward-open computed models, respectively (circled in **Figures 2C,D**). Hence, the ESM-AF2 protocol successfully modeled both the inward- and outward-open conformations of Znt8, as validated by comparison with experimental EC-derived contacts. A second test case for the ESM-AF2 modeling protocol using an SLC protein with both inward and outward-open experimental structures is presented for the *E. coli* D-galactonate:proton symporter (DgoT) in **Supplementary Figure S1**. Again the ESM-AF2 protocol successfully modeled both the inward- and outward-open conformations of DgoT, which could be validated by comparison with experimental EC-derived contacts

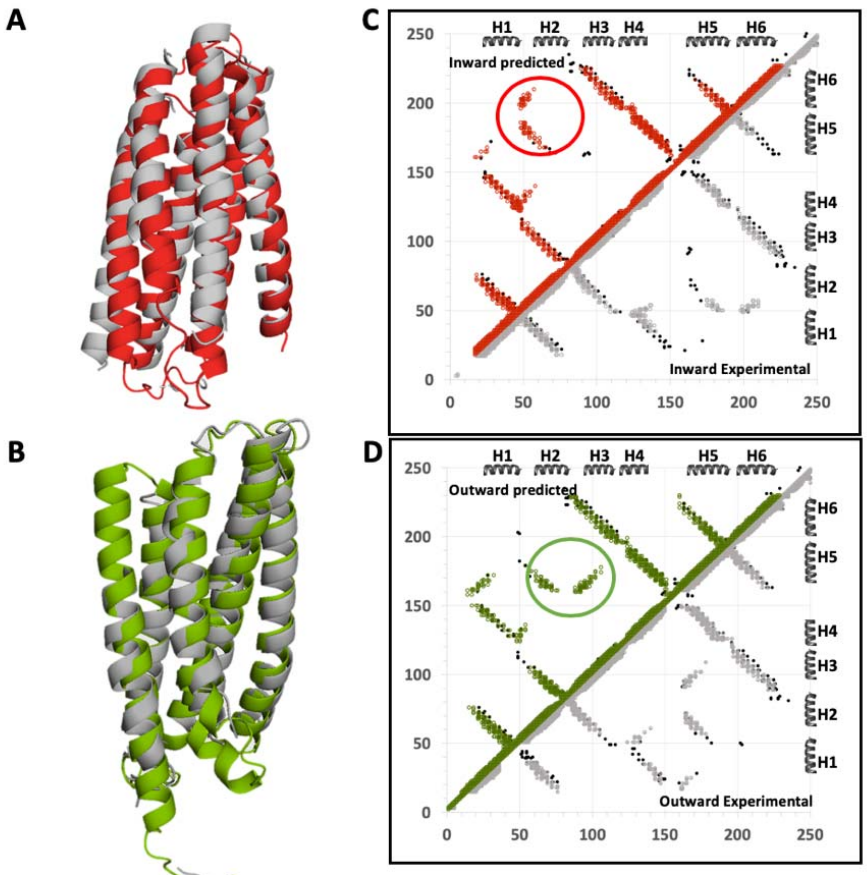


Fig. 2. Validation of ESM-AF2 protocol using and SLC protein with both outward- and inward-open experimental structures. The experimentally-determined cryo-EM structure of human ZnT8 WT in the absence of zinc has two chains, with one subunit in an inward-open conformation and the other in an outward-open conformation (PDB id: 6xpf chain A and B respectively). (A) Superposition of the AF2-predicted (red) and experimental (grey) inward-open structures, with backbone RMSD of 2.00 Å. (B) Superposition of outward-open model generated using the ESM-AF2 protocol (green) with the experimental (grey) outward-open structure, with backbone RMSD of 1.09 Å. (C) Comparison of the EC-

based contact map of ZnT8 (points shown in black) with contacts in the experimental (grey points) and predicted (red points) inward-open models. (D) Comparison of the EC-based contact map of ZnT8 (points shown in black) with contacts in the experimental (grey points) and predicted (green points) outward-open models. In panels C and D, major differences in the contact patterns of inward-open and outward-open states, supported by ECs unique to each state, are circled.

Modeling alternative conformations of SLC proteins for which AF modeling is dominated by the experimental structure of only one state. In the two cases above, we chose SLC proteins for which experimental structures of both outward- and inward-open conformations are available, and validated the ESM-AF2 modeling protocol against both the experimental atomic coordinates (using RMSD and GDT metrics) and against contact maps predicted from EC analysis which are based on experimental primary sequence data. However, for most SLC proteins, experimental structures are only available for one (or none) of the two states. We next modeled inward-open structures for two integral membrane proteins for which only the outward-open state is experimentally available. The results are shown in **Figure 3** for the 322-residue *Zea mays* CMP-sialic acid transporter 1 [PDB id 6i1r-A[55]], a SLC35A subfamily member and in **Figure 4** for the 337-residue *Saccharomyces cerevisiae* GDP-mannose sugar transporter 1 Vrg4 (PDB id 5oge[56]), an SLC35D subfamily member. For both proteins, only outward-open forms determined at 3.22 Å and 2.80 Å resolution, respectively, are available as X-ray crystal structures. In both of these cases, the ESM-AF2 protocol did not provide models of the inward-open state that could be validated by the EC-based contact map. However, using the ESM-MODELLER protocol, in which the outward-open state is modeled with AF2, and the inward-facing state is modeled using a “flipped-sequence” as input to *ESMfold*, providing template that is then used with a conventional template-based modeling approach, both outward- and inward-open states were generated. In both cases, the EC-based contact maps could be largely explained by the combined contact maps of these outward- and inward-open conformations, although some sporadic predicted ECs at the edge of the cutoff value used for identifying ECs were also present. These results validate the ESM-MODELLER process for cases where, due to the impact of memorization of conformational states available at the time of training on the AF inference, the ESM-AF2 method fails.

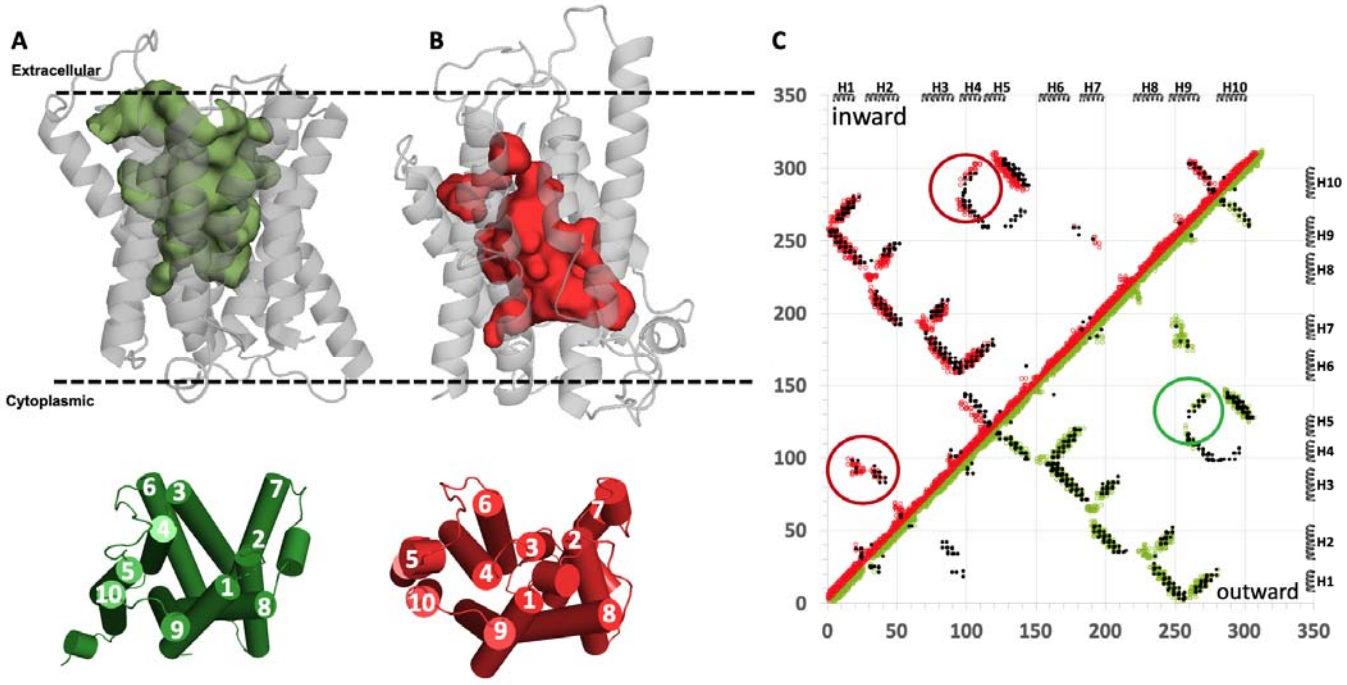


Fig. 3. ESM-MODELLER modeling of the inward-open conformation of the *Zea mays CMP-sialic acid transporter 1*. (A) The experimental outward-open structure (PDB id 6i1r-A). (B) The inward-open structure modeled using ESM-MODELLER. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) The combined contact maps of the two resulting models are consistent with the experimental EC-based contact map. Green contacts are those present in the experimental outward-open model, and red contacts are those present in the predicted inward-open model. EC-based contacts are shown as black dots. The EC-based contacts circled in green are unique to the outward-open conformation, and those circled in red are unique to the inward-open conformation. At the thresholds chosen for ECs several predicted contacts are not explained by the combination of two conformational states. In panels A and B (top), surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

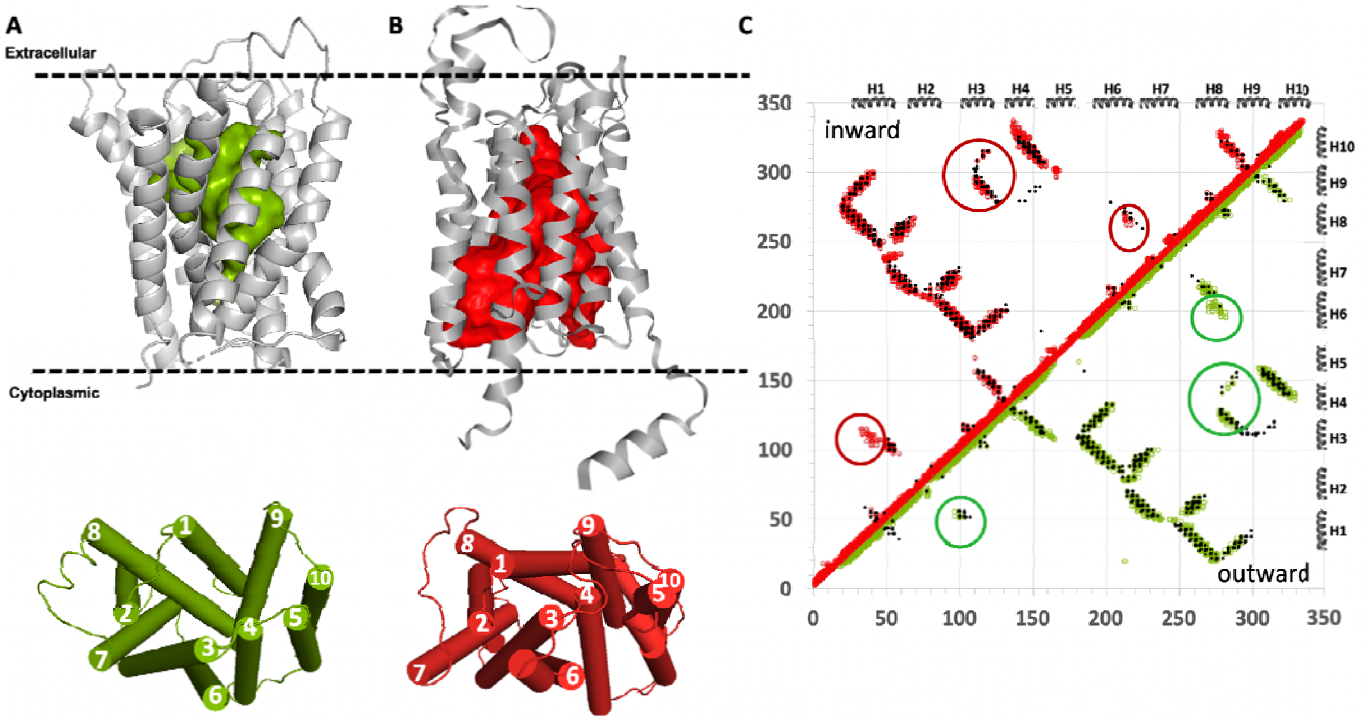


Fig. 4. ESM-MODELLER modeling of the inward-open conformation of the *S. cerevisiae* GDP-mannose sugar transporter 1, Vrg4. (A) The experimental outward-open structure (PDB id 5oge). (B) The inward-open structure modeled using ESM-AF2. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) The combined contact maps of the two resulting models are consistent with the EC-based contact map. EC-based contacts are shown as black dots, inward-open contacts as red circles and outward-open contacts as green circles. The EC-based contacts circled in green are unique to the outward-open conformation, and those circled in red are unique to the inward-open conformation. At the thresholds chosen for ECs several predicted contacts are not explained by the combination of two conformational states. In panels A and B, surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

Modeling alternative conformations of SLC35F2 with ESM-AF2. Of particular interest are SLC proteins for which no experimental structures are available for either the inward- or outward-facing states. SLC35F2 has < 12% sequence identity with the SLC35 subfamily members of known structure; in particular there is no good experimental structure that can be used as a template for comparative modeling of its inward- or outward-open conformations. Conventional AF2 modeling was carried out using the AF2-multimer colab server[51] executed both with the standard protocol without structural templates described in the Methods section and also with various other protocols using templates of distant homologues and multiple seeds. Modeling was also attempted using AF3[57]. Only the outward-open conformational state of SLC35 was returned by all permutations of AF2 and AF3 that were explored.

For SLC35F2 we further explored using shallow MSAs, dropouts, and the combination of dropouts with MSA masking to generate alternative conformational states. *AF-alt* was used to generate 480 models, and

AF_Sample and *AF_Sample2* were used to generate 3,000 models each. These enhanced sampling methods are very GPU intensive and require long run times. For this particular protein, for which no experimental structures were available in the PDB at the time of AF2 training, all three of these methods generated exclusively outward-open states (**Supplementary Figure S2A-C**). These results for SLC35F2 illustrate the common case where the currently available enhanced sampling methods fail to generate reliable models of multiple alternative conformational states. Interestingly, when *AF-Sample* was run on virtual flipped sequence of SLC35F2, exclusively inward-open conformational states were generated.

Having established the reliability, consistency, and limitations of the ESM-AF2 protocol, AF2 was used to model the outward-open conformation of SLC35F2, and ESM-AF2 was used to model its inward-open conformation (**Figure 5**). The contact maps of these two conformations were then compared with its EC-based contact map. The excellent agreement between the EC-based contact map and combined contact maps of the computed outward- and inward-open structures validate the accuracy of the ESM-AF2 protocol for modeling this conformational heterogeneity of SLC35F2.

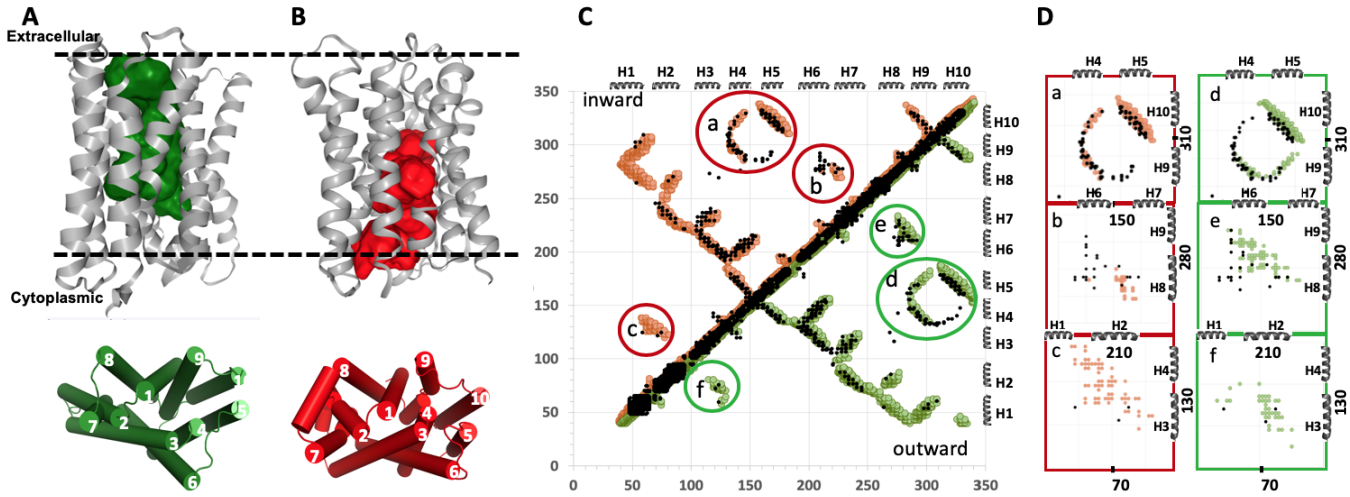


Fig. 5. AF2 / ESM-AF2 modeling of the outward- and inward-open conformations of human SLC35F2. (A) The outward-open structure modeled with AF2. (B) The inward-open structure modeled using ESM-AF2. In each of panels A and B the top images are ribbon representations of the protein structure with surface exposed cavities shown in either green (outward-open) or red (inward-open), and the bottom images are cylinder representations of these structural states with helices numbered 1 - 10. The dashed horizontal lines in panels A and B denote the approximate locations of the membrane boundaries. (C) Contact maps of the outward-open (green circles) and inward-open (red circles) structures superposed on the EC contact map (black dots). (D) Expanded regions (labeled a through f) of panel C, focusing on key distinguishing contacts and ECs between helices H4 and H10 in the inward-open conformation (subpanel D.a) and between helices H5 and H9 in the outward-open conformation (subpanel D.d). Also shown in panel D, subpanels a, b, c, e and f are other key contacts and ECs distinguishing the two states. In panels A and B (top), surface pockets are represented as space-filled voids using the server https://kvfinder-web.cnpem.br/.

Modeling alternative conformational states of other SLC proteins. We selected 4 additional SLC proteins for modeling with the ESM-AF2 or ESM-MODELLER protocol. These results are summarized

in **Supplementary Table S1** and accompanying **Supplementary Figures S3 – S6**. In all of these cases for which structures of one conformational state were available in the PDB at the time of AF2 training, bias toward this state was observed when using AF2 alone or even when using AF2 with a template for the alternative state generated with ESM using a flipped sequence; i.e. the ESM-AF2 protocols described here fail to identify the alternative conformational state when one conformational state was available in the PDB at the time of AF2 training. However, the ESM-MODELLER protocol, which avoids the bias of conformational state modeling due to "memorization" often observed using AF2, provided models of both inward-facing and outward-facing states, with excellent agreement (< 1 - 2 Å rmsd) to experimental models where available, and in concordance with EC-predicted contact maps.

DISCUSSION

We have developed and tested hybrid AF2 / ESM-AF2 and AF2 / ESM-MODELLER protocols for modeling alternative conformations of pseudo-symmetric SLC transporters. Generally, where conventional AF2 modeling (or even AF2 modeling with enhanced sampling) provides only one (either inward- or outward-open) conformational state; the alternative state can then be modeled by the ESM-AF2 (or ESM-MODELLER) protocol. The ESM-AF2 protocol is inspired by a more traditional approach using comparative modeling of the pseudo-symmetric halves of SLC transporters[28,41-47]. This traditional approach requires an accurate sequence alignment between the two symmetric halves of SLC protein. However, in some cases it is difficult to determine the correct sequence alignment needed for accurate comparative modeling. In the ESM-AF2 (or ESM-MODELLER) approach, we use *ESMfold* to generate from a virtual flipped sequence a virtual protein structure, which is then used as a structure modeling template without the need for any sequence alignment between the two halves of the SLC protein. This allowed us to reliably model alternative conformational states of several SLC transporters that were difficult to model using the traditional approach. The resulting multi-state models are validated by comparison with sequence-based evolutionary co-variance data (ECs) that encode information about contacts present in the various conformational states adopted by the protein.

The ESM-AF2 (or ESM-MODELLER) approach is simple to implement and runs fast using publicly-available servers. However, despite the successful examples demonstrated in this study, these ESM-AF2 (or ESM-MODELLER) protocols for modeling alternative conformational states of pseudo-symmetric SLC proteins have some limitations. In particular, where structures of only one of the alternative states was available in the PDB at the time of AF2 training, a significant bias towards this state was observed when AF2 was used either directly or as part of the ESM-AF2 modeling process. Another shortcoming is that they cannot be applied directly to homodimeric pseudo-symmetric SLC proteins, such as YiiP or EmrE[24,58]. Coordinates of SLC proteins with large loops and other structural decorations require manual editing to eliminate these loops / decorations prior to applying the protocol. In addition, the validation of alternative state conformations by contact predictions relies on the quality of these contact predictions, and may not work well for SLC sequence families for which only shallow MSAs are available.

The ability of AF2 to model protein structures not included in its training has been demonstrated in various CASP blind assessments[59,60]. Conventional AF2 was also reported to be successful in accurate modeling of protein structures determined by NMR methods which were not included in its training data, and for which no structures of homologous proteins were available at the time of training[61]. However, for proteins adopting multiple conformational states, the training carried out with training data that includes only one state may bias the predictor and limit its ability to model the alternative state. Recently Porter and co-workers have demonstrated that, at least for fold-flipping proteins which have significant structural differences between conformational states, AF2 modeling with enhanced sampling is often biased toward the conformational state reported in the PDB and potentially used in the AF2 training, and often is not able to predict conformational states not represented in the AF2 training data. Combining >280,000 models from several implementations of AF2 and AF3, a 35% success rate was achieved for fold switchers likely in AF's training sets[9]. Bryant and Noé also have explored this question by training a structure prediction network, *Cfold*, on a conformational split of the PDB that excludes alternative conformations for protein structure pairs solved in two conformational states. While > 50% of experimentally-known nonredundant alternative protein conformations evaluated were predicted with high accuracy (TM-score > 0.8), for the remaining pairs *Cfold* failed to correctly model the alternative conformational state not included in the training data[11]. These results indicate that while in some cases, the network has learned enough to model alternative conformational states not included in the training data[62], in other cases success may in fact rely on some kind of memorization; i.e. both factors can be at play. It has also been suggested that *ESMfold* may be less sensitive to this memorization bias[10]. Consistent with these observations, we also observed a bias toward previously reported conformational states when modeling with either AF2 or with the ESM-AF2 modeling protocol outlined here. For these SLC proteins, this bias is overcome using the ESM-MODELLER protocol. Though limited to the specific but important class of pseudo-symmetric SLC proteins, the ESM-MODELLER protocol can be used routinely for the many proteins in this class to generate models for both conformational states regardless of memorization bias related to the training data. However, where no memorization bias is involved, the ESM-AF2 protocol is preferable as template-guided AF2 has more accurate properties than conventional template-based modeling method. More significantly, the fact that multistate models of SCL35F2 consistent with EC-based contact maps could be generated using ESM-AF2 when no structural templates are available shows that it is possible to suppress this conformational bias, and suggests a general protocol using a retrained AF network in which all homologous structures which might bias the inference to specific conformational states are excluded from the training data.

Although we have focused our analysis on the outward and inward conformational states of SLC transporters, intermediate "occluded" states have also been captured in X-ray crystal and cryoEM structures. Although the ESM-AF2/MODELLER protocol could potentially also generate such occluded states, this was not observed in the cases studied here.

Conclusions. In this work we describe, validate, and compare hybrid ESM-AF2 and ESM-MODELLER protocols for modeling alternative conformational states of pseudo-symmetric SLC proteins. The

approach overcomes the shortcoming of conventional AF2 structure calculations which generally provide only one of the multiple conformational states observed experimentally. The method is simple to use, rapid to run, and can be implemented using the public domain servers. In this approach, the resulting multi-state models are validated by comparison with sequence-based EC data that encode information about contacts present in the various conformational states adopted by the protein. Overall, the current study validates the ESM-AF2/MODELLER protocol for modeling conformational heterogeneity of pseudo-symmetric SLC transporters, one of the most extensive class of transporters in the human proteome.

ACKNOWLEDGEMENTS

We thank Dr. Davide Sala for providing scripts for running *AF-alt*, T.B. Acton, T. Benavides, A. De Falco, K. Fraga, A. Gaur, R. Greene-Cramer, Y.J. Huang, T.A. Ramelot, B. Shurina, L. Spaman, and R. Tejero for helpful discussions and comments on the manuscript, and S. Collen for computer system administration support. This work was supported financially by National Institutes of Health NIGMS grants R35 GM141818 (to G.T.M.) and R35 GM122518 (to M.J.R.), and by the Rensselaer Polytechnic Institute (RPI) Bio-computing and Bio-informatics Constellation Chair Fund. GTM also acknowledges access to the RPI Center for Computational Innovations (CCI) computing infrastructure.

AUTHOR CONTRIBUTIONS

GVTS, ND, MJR, and GTM jointly conceptualized the study and analyzed data. GVTS carried out bioinformatics analyses and generated graphics. All authors contributed in writing and editing the manuscript.

DECLARATION OF INTERESTS

GTM is a founder of Nexomics Biosciences, Inc. This does not represent a conflict of interest for this study.

SUPPLEMENTARY MATERIAL

Table S1. Examples of modeling of both outward-open and inward-open states of SLC proteins.

Fig. S1. Validation of ESM-AF2 protocol using an SLC protein with both outward- and inward-open experimental structures.
Fig. S2. Conformational states of SLC35F2 were modeled using various enhanced sampling protocols.
Fig. S3. Thiamine transporter 1 (SLC19A2).
Fig. S4. Aromatic amino acid exporter YddG.
Fig. S5. Reduced folate transporter (SLC19A1).

REFERENCES

1 Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

2 Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379, 1123-1130 (2023). https://doi.org:10.1126/science.ade2574

3 Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science 373, 871-876 (2021). https://doi.org:10.1126/science.abj8754

4 Ahdritz, G. et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nature Methods (2024). https://doi.org:10.1038/s41592-024-02272-z

5 Huang, Y. J. et al. Assessment of prediction methods for protein structures determined by NMR in CASP14: impact of AlphaFold2. PROTEINS: Structure, Function and Bioinformatics (2021).

6 Del Alamo, D., Sala, D., McHaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. Elife 11 (2022). https://doi.org:10.7554/eLife.75751

7 Saldano, T. et al. Impact of protein conformational diversity on AlphaFold predictions. Bioinformatics 38, 2742-2748 (2022). https://doi.org:10.1093/bioinformatics/btac202

8 Wayment-Steele, H. K. et al. Predicting multiple conformations via sequence clustering and AlphaFold2. Nature 625, 832-839 (2024). https://doi.org:10.1038/s41586-023-06832-9

9 Chakravarty, D. et al. AlphaFold predictions of fold-switched conformations are driven by structure memorization. Nat Commun 15, 7296 (2024). https://doi.org:10.1038/s41467-024-51801-z

10 Xie, T. & Huang, J. Can protein structure prediction methods capture alternative conformations of membrane transporters? J Chem Inf Model 64, 3524-3536 (2024). https://doi.org:10.1021/acs.jcim.3c01936

11 Bryant, P. & Noé, F. Structure prediction of alternative protein conformations. Nature Communications 15, 7328 (2024). https://doi.org:10.1038/s41467-024-51507-2

12 Kalakoti, Y. & Wallner, B. AFsample2: Predicting multiple conformations and ensembles with AlphaFold2. bioRxiv, 2024.2005.2028.596195 (2024). https://doi.org:10.1101/2024.05.28.596195

13 Sala, D., Engelberger, F., McHaourab, H. S. & Meiler, J. Modeling conformational states of proteins with AlphaFold. Curr Opin Struct Biol 81, 102645 (2023). https://doi.org:10.1016/j.sbi.2023.102645

14 Heo, L. & Feig, M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. PROTEINS: Structure, Function and Bioinformatics 90, 1873-1885 (2022). https://doi.org:10.1002/prot.26382

15 Sala, D., Hildebrand, P. W. & Meiler, J. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. Front Mol Biosci 10, 1121962 (2023). https://doi.org:10.3389/fmolb.2023.1121962

16 Porter, L. L., Chakravarty, D., Schafer, J. W. & Chen, E. A. ColabFold predicts alternative protein structures from single sequences, coevolution unnecessary for AF-cluster. bioRxiv, 2023.2011.2021.567977 (2023). https://doi.org:10.1101/2023.11.21.567977

17 Stein, R. A. & McHaourab, H. S. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. PLOS Computational Biology 18, e1010483 (2022). https://doi.org:10.1371/journal.pcbi.1010483
18 Johansson-Akhe, I. & Wallner, B. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. Front Bioinform 2, 959160 (2022). https://doi.org:10.3389/fbinf.2022.959160
19 Wallner, B. AFsample: improving multimer prediction with AlphaFold using massive sampling. Bioinformatics 39 (2023). https://doi.org:10.1093/bioinformatics/btad573
20 Wallner, B. Improved multimer prediction using massive sampling with AlphaFold in CASP15. Proteins 91, 1734-1746 (2023). https://doi.org:10.1002/prot.26562
21 Benavides, T. L. & Montelione, G. T. Integrative Modeling of Protein-Polypeptide Complexes by Bayesian Model Selection using AlphaFold and NMR Chemical Shift Perturbation Data. bioRxiv (2024). https://doi.org:10.1101/2024.09.19.613999
22 Hediger, M. A., Clemencon, B., Burrier, R. E. & Bruford, E. A. The ABCs of membrane transporters in health and disease (SLC series): introduction. Mol Aspects Med 34, 95-107 (2013). https://doi.org:10.1016/j.mam.2012.12.009
23 Colas, C., Ung, P. M. & Schlessinger, A. SLC transporters: Structure, function, and drug discovery. Medchem Comm 7, 1069-1081 (2016). https://doi.org:10.1039/C6MD00005C
24 Bai, X., Moraes, T. F. & Reithmeier, R. A. F. Structural biology of solute carrier (SLC) membrane transport proteins. Mol Membr Biol 34, 1-32 (2017). https://doi.org:10.1080/09687688.2018.1448123
25 Pizzagalli, M. D., Bensimon, A. & Superti-Furga, G. A guide to plasma membrane solute carrier proteins. FEBS J 288, 2784-2835 (2021). https://doi.org:10.1111/febs.15531
26 Sarangi, A., Bupp, K. & Roth, M. J. Identification of a retroviral receptor used by an envelope protein derived by peptide library screening. Proc Natl Acad Sci U S A 104, 11032-11037 (2007). https://doi.org:10.1073/pnas.0704182104
27 Fredriksson, R., Nordstrom, K. J., Stephansson, O., Hagglund, M. G. & Schioth, H. B. The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. FEBS Lett 582, 3811-3816 (2008). https://doi.org:10.1016/j.febslet.2008.10.016
28 Forrest, L. R. Structural biology. (Pseudo-)symmetrical transport. Science 339, 399-401 (2013). https://doi.org:10.1126/science.1228465
29 Leano, J. B. et al. Structures suggest a mechanism for energy coupling by a family of organic anion transporters. PLoS Biol 17, e3000260 (2019). https://doi.org:10.1371/journal.pbio.3000260
30 Xue, J., Xie, T., Zeng, W., Jiang, Y. & Bai, X. C. Cryo-EM structures of human ZnT8 in both outward- and inward-facing conformations. Elife 9 (2020). https://doi.org:10.7554/eLife.58823
31 Wang, N. et al. Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. Cell 184, 370-383.e313 (2021). https://doi.org:10.1016/j.cell.2020.11.043
32 Killer, M., Wald, J., Pieprzyk, J., Marlovits, T. C. & Löw, C. Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes. Science Advances 7, eabk3259 (2021). https://doi.org:doi:10.1126/sciadv.abk3259
33 Lu, Y. et al. Structural insights into the conformational changes of BTR1/SLC4A11 in complex with PIP2. Nature Communications 14, 6157 (2023). https://doi.org:10.1038/s41467-023-41924-0

34 Hopf, T. A. et al. Three-dimensional structures of membrane proteins from genomic sequencing. Cell 149, 1607-1621 (2012). https://doi.org:10.1016/j.cell.2012.04.012
35 Morcos, F., Jana, B., Hwa, T. & Onuchic, J. N. Coevolutionary signals across protein lineages help capture multiple protein conformations. Proc Natl Acad Sci U S A 110, 20533-20538 (2013). https://doi.org:10.1073/pnas.1315625110
36 Sutto, L., Marsili, S., Valencia, A. & Gervasio, F. L. From residue coevolution to protein conformational ensembles and functional dynamics. Proc Natl Acad Sci U S A 112, 13567-13572 (2015). https://doi.org:10.1073/pnas.1508584112
37 Toth-Petroczy, A. et al. Structured States of Disordered Proteins from Genomic Sequences. Cell 167, 158-170 e112 (2016). https://doi.org:10.1016/j.cell.2016.09.010
38 Zheng, S. et al. Structure and mutagenic analysis of the lipid II flippase MurJ from Escherichia coli. Proc Natl Acad Sci U S A 115, 6709-6714 (2018). https://doi.org:10.1073/pnas.1802192115
39 Huang, Y. J. et al. Combining Evolutionary Covariance and NMR Data for Protein Structure Determination. Methods Enzymol 614, 363-392 (2019). https://doi.org:10.1016/bs.mie.2018.11.004
40 Schafer, J. W. & Porter, L. L. Evolutionary selection of proteins with two folds. Nat Commun 14, 5478 (2023). https://doi.org:10.1038/s41467-023-41237-2
41 Crisman, T. J., Qu, S., Kanner, B. I. & Forrest, L. R. Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. Proc Natl Acad Sci U S A 106, 20752-20757 (2009). https://doi.org:10.1073/pnas.0908570106
42 Radestock, S. & Forrest, L. R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. J Mol Biol 407, 698-715 (2011). https://doi.org:10.1016/j.jmb.2011.02.008
43 Schushan, M. et al. A model-structure of a periplasm-facing state of the NhaA antiporter suggests the molecular underpinnings of pH-induced conformational changes. J Biol Chem 287, 18249-18261 (2012). https://doi.org:10.1074/jbc.M111.336446
44 Liao, J. et al. Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger. Science 335, 686-690 (2012). https://doi.org:10.1126/science.1215759
45 Kowalczyk, L. et al. Molecular basis of substrate-induced permeation by an amino acid antiporter. Proc Natl Acad Sci U S A 108, 3935-3940 (2011). https://doi.org:10.1073/pnas.1018081108
46 Mancusso, R., Gregorio, G. G., Liu, Q. & Wang, D. N. Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. Nature 491, 622-626 (2012). https://doi.org:10.1038/nature11542
47 Kim, J. et al. Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. Nature 576, 315-320 (2019). https://doi.org:10.1038/s41586-019-1795-x
48 Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics 54, 5.6.1-5.6.37 (2016). https://doi.org:10.1002/cpbi.3
49 He, B., Mortuza, S. M., Wang, Y., Shen, H. B. & Zhang, Y. NeBcon: protein contact map prediction using neural network training coupled with naïve Bayes classifiers. Bioinformatics 33, 2296-2306 (2017). https://doi.org:10.1093/bioinformatics/btx164
50 Vehlow, C. et al. CMView: interactive contact map visualization and analysis. Bioinformatics 27, 1573-1574 (2011). https://doi.org:10.1093/bioinformatics/btr163

51 Mirdita, M. et al. ColabFold: making protein folding accessible to all. Nat Methods 19, 679-682 (2022). https://doi.org:10.1038/s41592-022-01488-1
52 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815 (1993). https://doi.org:10.1006/jmbi.1993.1626
53 Zemla, A. et al. AS2TS system for protein structure modeling and analysis. Nucleic Acids Res 33, W111-115 (2005). https://doi.org:10.1093/nar/gki457
54 Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, W296-w303 (2018). https://doi.org:10.1093/nar/gky427
55 Nji, E., Gulati, A., Qureshi, A. A., Coincon, M. & Drew, D. Structural basis for the delivery of activated sialic acid into Golgi for sialyation. Nat Struct Mol Biol 26, 415-423 (2019). https://doi.org:10.1038/s41594-019-0225-y
56 Parker, J. L. & Newstead, S. Structural basis of nucleotide sugar transport across the Golgi membrane. Nature 551, 521-524 (2017). https://doi.org:10.1038/nature24464
57 Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493-500 (2024). https://doi.org:10.1038/s41586-024-07487-w
58 Fleishman, S. J. et al. Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. J Mol Biol 364, 54-67 (2006). https://doi.org:10.1016/j.jmb.2006.08.072
59 Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. Proteins 89, 1607-1617 (2021). https://doi.org:10.1002/prot.26237
60 Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XV. Proteins 91, 1539-1549 (2023). https://doi.org:10.1002/prot.26617
61 Li, E. H. et al. Blind assessment of monomeric AlphaFold2 protein structure models with experimental NMR data. J Magn Reson 352, 107481 (2023). https://doi.org:10.1016/j.jmr.2023.107481
62 Roney, J. P. & Ovchinnikov, S. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. Phys Rev Lett 129, 238101 (2022). https://doi.org:10.1103/PhysRevLett.129.238101

SUPPLEMENTARY INFORMATION

Supplementary Table S1. Examples of modeling of both outward-open and inward-open states of pseudo-symmetric integral membrane proteins.

Name	Uniprot id	Experimental structure (O: Outward open) (I: Inward open)	AF2 prediction (rmsd Å)	AF Database (rmsd Å)	ESM-AF (rmsd Å)	ESM-MODELLER (rmsd Å)	
E. coli D-galactonate:proton symporter SLC17 DgoT	P0AA76	6e9o (O) 6e9n (I)	0.76 -	0.85	1.35	3.8	**Supplementary Figure S2**
Thiamine transporter 1 SLC19A2	O60779	n/a (O) n/a (I)	- modeled	n/a n/a	modeled -	modeled -	**Supplementary Figure S3**
Aromatic amino acid exporter YddG	D7A5Q8	5i20 (O) n/a (I)	0.68 -	0.43	0.45	- modeled	**Supplementary Figure S4**
Reduced folate transporter SLC19A1	P41440	n/a (O) 7xtk (I)	- 0.64	0.58	0.59	modeled -	**Supplementary Figure S5**
Chloroquine resistance transporter I	W7FI62	6ukj (O) n/a (I)	0.89 -	2.76	0.77	- modeled	**Supplementary Figure S6**
Human proton-coupled zinc antiporter SLC30A8 ZnT8	Q8IWU4	6xpf_B (O) 6xpf_A (I)	1.09 -	1.12	2.00		**Figure 2B**
Zea mays CMP-sialic acid transporter 1 SLC35A1	B4FZ94	6i1r (O) n/a (I)	0.69 -	0.26	0.65	- modeled	**Figure 3**
S. cerevisiae GDP-mannose sugar transporter 1, SLC35D Vrg4.	P40107	5oge (O) n/a (I)	0.56 -	0.46	0.52	- modeled	**Figure 4**

For each SLC protein, conventional AF2 modeling provided either an inward-open (I) or outward-open (O) state for which the backbone root-mean-squared deviation to an available X-ray crystal or cryoEM structure is listed. The alternative outward-open or inward-open conformation was then generated with the ESM-AF2 or ESM-MODELLER protocol. For these models of alternative conformational states, backbone rmsd's are reported where experimental structures are available. Models for each protein available in the AlphaFold2 data base are indicated along with backbone rmsd's to the most similar experimental structure. n/a – not available.

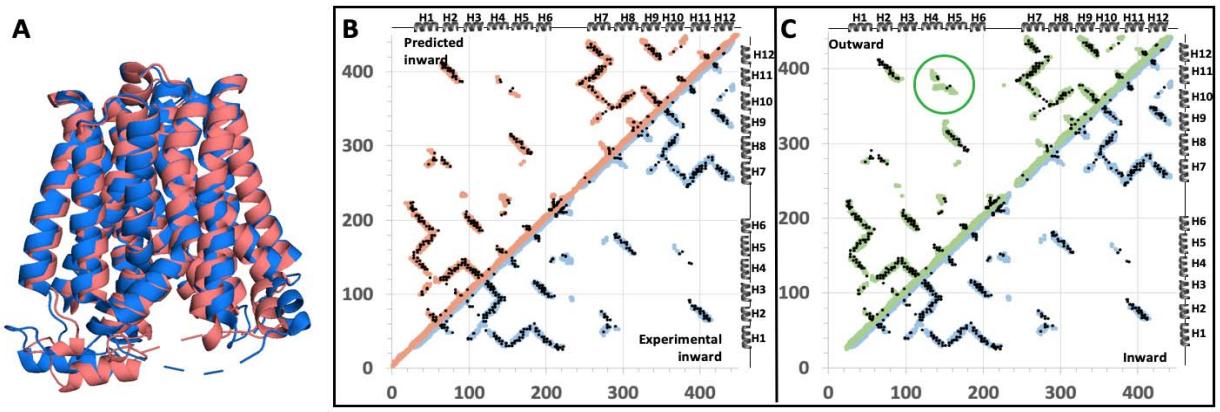


Fig. S1. Validation of ESM-AF2 protocol using an SLC protein with both outward- and inward-open experimental structures. (A) Superposition of the inward-open conformational state of *E. coli* D-galactonate:proton symporter (SLC17, DgoT) modeled using the ESM-AF2 protocol from a template generated by *ESMfold* using a virtual flipped-sequence (red) with the experimentally-determined inward-open structure PDB id 6e9n (blue). (B) Contact maps for the inward-open conformational state of DgoT predicted by ESM-AF2 (red circles), observed in the cryoEM experimental structure (blue circles), and predicted by EC analysis (black dots). (C) Contact maps for the experimentally-determined outward-open conformational state of DgoT (green circles ; PDB id 6e9o) and experimentally-determined inward-open conformational state (blue circles; PDB id 6e9n) structures compared with contact predictions from EC analysis (black dots). In this example, only a few EC-based contacts between helices H4 and H11 and helices H5 and H10 (circled in green) distinguish the inward-open from outward-open structures.

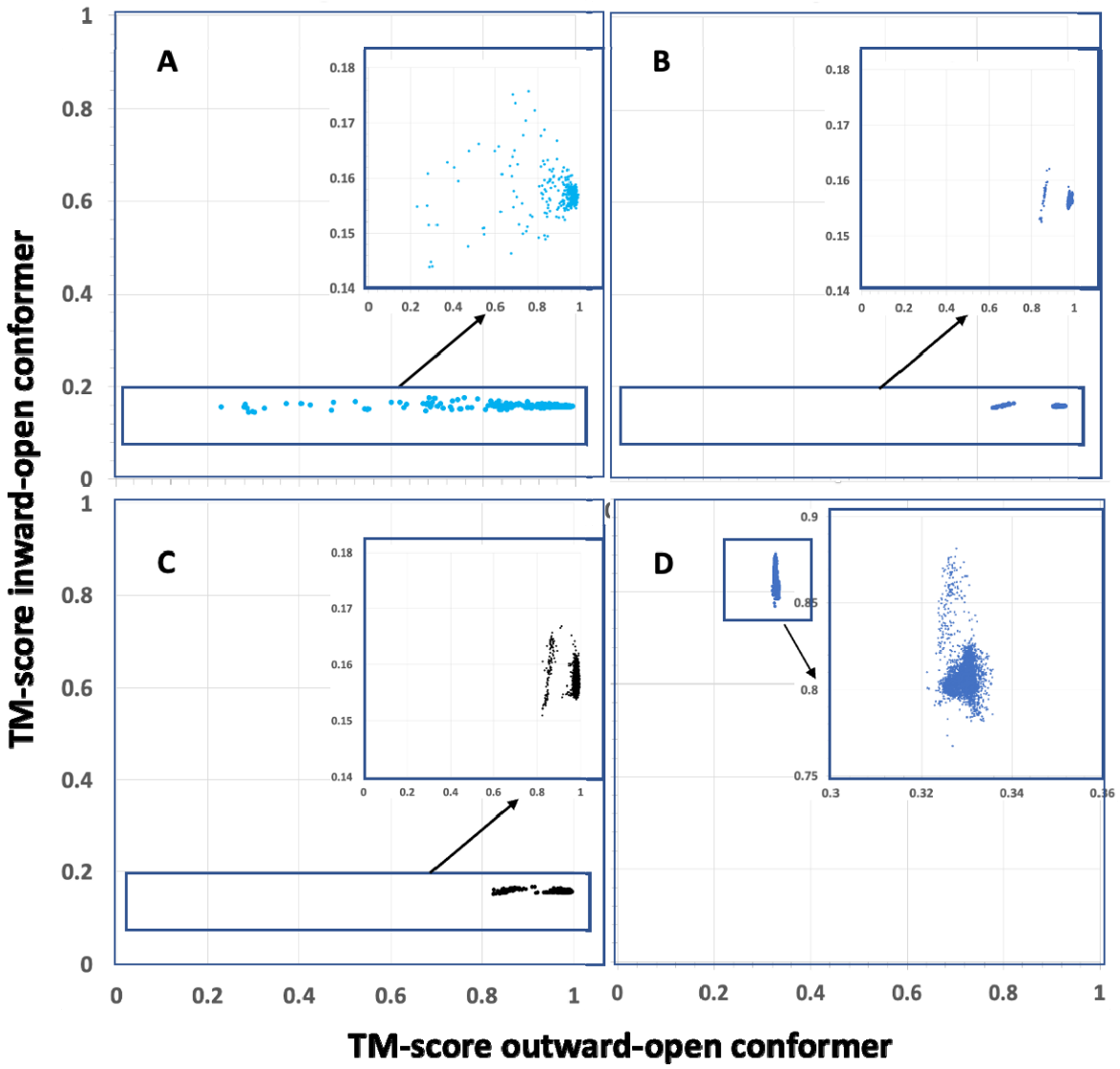


Fig. S2. Conformational states of SLC35F2 were modeled using various enhanced or massive sampling protocols as implemented in (A) *AlphaFold-alt* (480 models), (B) *AlphaFold-sample* (3,000 models), and (C) *AlphaFold-sample2* (3,000 models); the resulting models were compared with the outward-open and inward-open models from the standard AF2 (for outward-facing) and our ESM-AF2 protocol (for inward-facing). None of these three methods modeled the alternative "inward-open" conformational state. (D) Massive sampling with *AF-sample2* (3,000 models) using a "flipped" protein sequence generated only inward-open conformations for SLC35F2 can provide an inward-facing structure of SLC35F2. In each panel, the inset shows an expanded region of the plot. In all cases, no models were observed outside of the boxed regions. Only models with pLDDT > 70 are plotted.

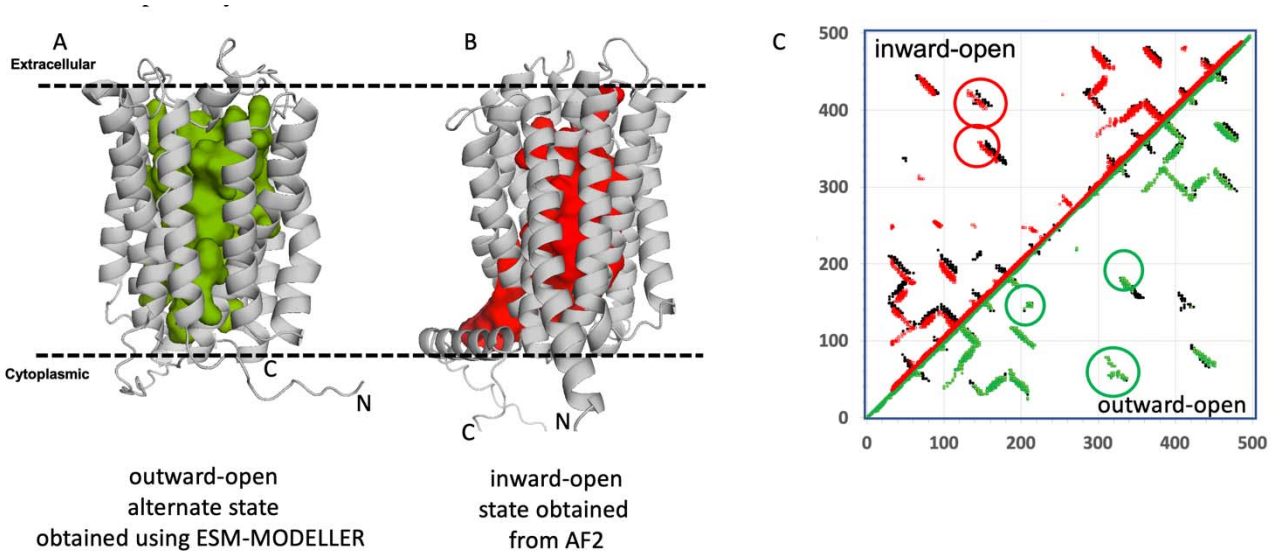


Fig. S3. Thiamine transporter 1 (SLC19A2). (A) The outward-open conformer generated using the ESM-MODELLER method. (B) AF2 models the inward-open state. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

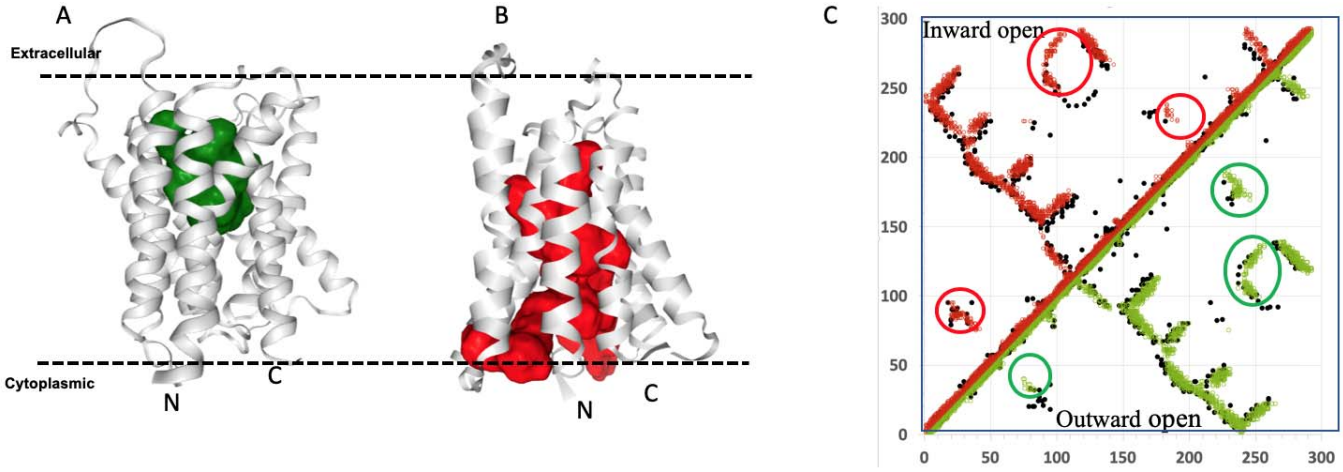


Fig. S4. Aromatic amino acid exporter YddG. (A) AF2 models the outward-open state. (B) The inward-open conformer was generated using the ESM-MODELLER method. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

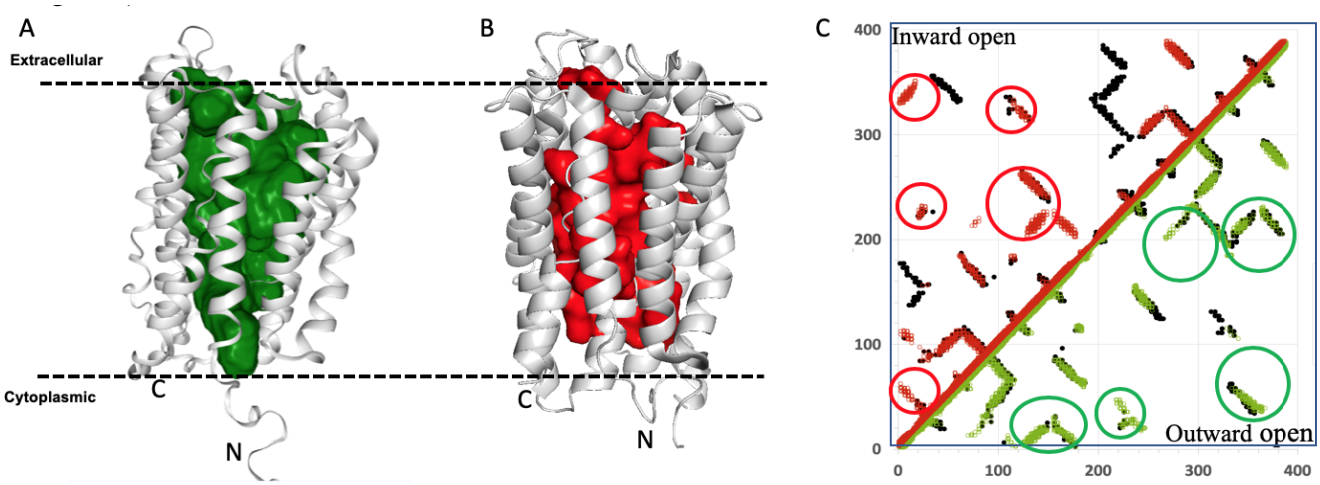


Fig. S5. Reduced folate transporter (SLC19A1). (A) The outward-open conformer was generated using the ESM-MODELLER method. (B) AF2 models the inward-open state. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.

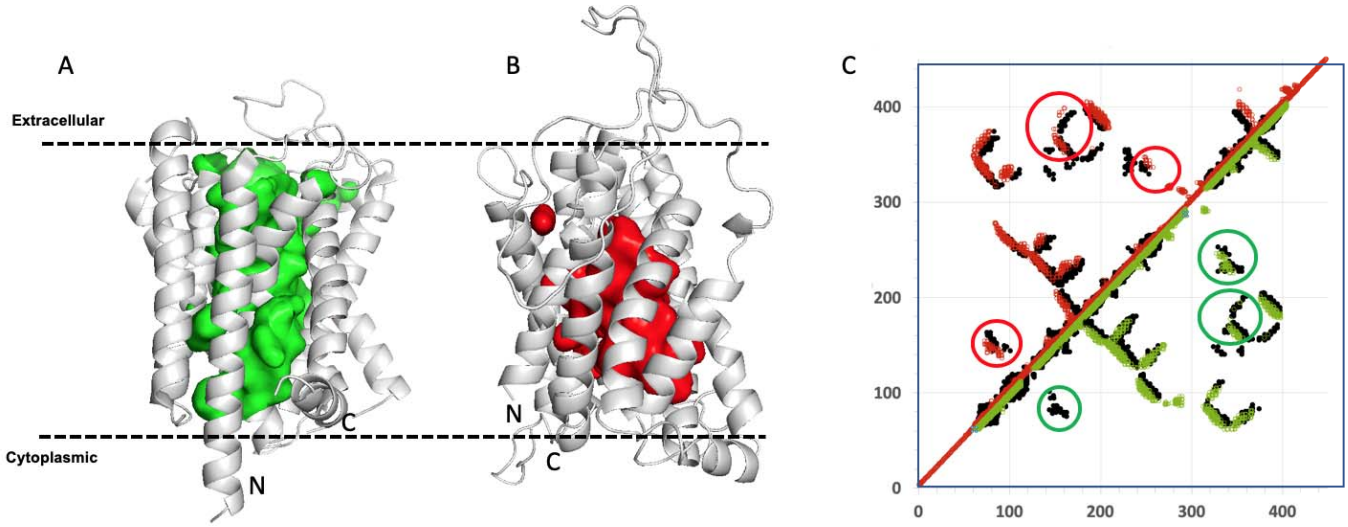


Fig. S6. Chloroquine resistance transporter I. (A) AF2 models the outward-open state. (B) The inward-open conformer was generated using the ESM-MODELLER method. (C) The outward-open conformer contact map (green, lower diagonal), and inward-open conformer contact map (red, upper diagonal), generated using Cα-Cα cutoff of 10 are superposed on the symmetric EC-predicted contact map (black). ECs unique to each state are indicated with green and red circles, respectively.